

DOCUMENT RESUME

ED 301 592

TM 012 535

AUTHOR Patience, Wayne; Auchter, Joan
TITLE Establishing and Maintaining Score Scale Stability and Reading Reliability.
PUB DATE Apr 88
NOTE 40p.; Paper presented at the Annual Meeting of the National Testing Network in Writing (Minneapolis, MN, April 15-17, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Equivalency Tests; *Essay Tests; *Scoring; Secondary Education; Testing; *Test Reliability; *Writing Evaluation
IDENTIFIERS *General Educational Development Tests; Reading Reliability; Score Stability; *Writing Skills Test

ABSTRACT

A central aim in any assessment program is to ensure fair and stable scoring from administration to administration. When administrations are decentralized, not only in location, but in frequency and in logistical configuration, it is imperative to construct training, certifying, and monitoring systems that provide continuity between the original application of scoring standards and the application of those standards from testing site to testing site and from test administration to test administration. The 1988 Tests of General Educational Development (GED Tests) include a direct measure of writing as part of the Writing Skills Test (WST). Two reliability concerns influenced the choice of methods used to incorporate an essay into the GED WST: (1) scale stability; and (2) reading reliability. Much of the GED Testing Service's research on the feasibility of including an essay focused on defining and maintaining an holistic score scale and methods of ensuring comparable reading reliability across scoring sessions and decentralized sites. Because these scores, along with those on the other four subject area tests in the battery, determine whether an examinee receives a high school equivalency credential, reliability of the WST is of major importance. Descriptions of the GED tests, the WST, and the scoring method and guide; a sample GED essay test; seven data tables; two graphs; and selected scoring formulas are appended. (Author/TJH)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED301592

Establishing and Maintaining Score Scale Stability and Reading Reliability

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

Wayne Patience and Joan Auchter
GED Testing Service

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

WAYNE M. PATIENCE

- ✓ This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Abstract

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

A central aim in any assessment program is to ensure fair and stable scoring from administration to administration. When administrations are decentralized, not only in location, but in frequency and in logistical configuration, it is imperative to construct training, certifying, and monitoring systems which provide continuity between the original application of scoring standards and the application of those standards from testing site to testing site and from test administration to test administration. The 1988 Tests of General Educational Development (GED Tests) include a direct measure of writing as part of the Writing Skills Test. The following reliability concerns influenced the choice of methods used to incorporate an essay into the GED Writing Skills Test: scale stability--the degree to which readers uniformly apply a defined score scale in evaluating essays both within and across readings (scoring sessions); reading reliability--the consistency with which different readers in a given scoring session award the same scores to a given set of essays. Much of the GED Testing Service's research on the feasibility of including an essay focused on defining and maintaining an holistic score scale and methods of ensuring comparable reading reliability across scoring sessions and decentralized sites. Because these scores, along with those on the other four subject area tests in the battery, determine whether an examinee receives a high school equivalency credential, reliability of the Writing Skills Test is of major importance.

Paper presented at the annual meeting of The National Testing Network in Writing, Minneapolis, Minnesota, April, 1988.

TM012 535

Establishing and Maintaining Score Scale Stability and Reading Reliability

Wayne Patience and Joan Auchter

The primary purpose of this presentation is to describe and illustrate the methods employed by GEDTS to establish and maintain score stability and reading reliability. More specifically, the focus is on addressing reader and site certification procedures, as well as scoring site monitoring techniques. But before turning to this focus, some background is required to understand the context in which choices were made in developing the 1988 GED Writing Skills Test.

Purpose of the Tests of General Educational Development (GED Tests)

GED Tests are used extensively by states, Canadian provinces, and U. S. and Canadian territories as the basis for the award of a high school equivalency diploma to those who did not complete a high school program. Nearly three-quarters of a million examinees are tested annually, and almost half a million people earn a GED credential each year. Approximately one in seven high school completion credentials issued in the United States is a GED credential.

The credential issued on the basis of the GED Test results is commonly referred to as a high school equivalency diploma. This notion of equivalency derives from: 1) defining the content of the GED Tests so as to reflect the commonly expected outcomes of completing a traditional high school program of study, and 2) defining passing scores relative to the actual demonstrated performance of contemporary graduating seniors.

The authority to award a high school equivalency diploma rests with the participating jurisdictions. In part, this explains the extremely decentralized nature of the GED Program. GED Tests are administered in nearly 3,000 Official GED Testing Centers across North America, and are given nearly every day of the year. Appendix A provides a more detailed overview of the GED Program and Tests.

Adding an Essay to the GED Writing Skills Test

Periodically, the GED Tests are extensively reviewed to evaluate how well they reflect current high school curriculum and expectations. Such a review was begun in 1983 by surveying secondary school curriculum specialists, adult education teachers and program directors, and GED Administrators and Examiners (Spille, 1984). National secondary school associations were also polled, and, in addition to being asked to complete the surveys, they were asked to provide nominations for individuals specializing in the major secondary school subject areas (English, Social Studies, Science, and Mathematics) who would be well qualified to serve on a GED Tests Specifications Committee. From these nominations, Committee members were selected, and, in February, 1984, the Committee met after having reviewed the current GED Tests and many reports regarding secondary school curriculum. In June of 1984, the Committee released its GED Tests Specifications Report (GEDTS Tests Specification Committee, 1984). Among the primary recommendations of the

Committee was that the GED Testing Service add an essay to the GED Writing Skills Test.

After reviewing the research on large scale writing assessment and conducting research on: 1) the relationship of the current multiple-choice Writing Skills Test and direct measures of writing skills (GED Testing Service, 1985a), and 2) reliability and validity issues of using a direct measure of writing as part of the GED Tests (GEDTS, 1985b), the GED Testing Service staff recommended the incorporation of an essay to the Commission on Educational Credit and Credentials, the independent policy making body for the GED Testing Service. With attention to the views of participating jurisdictions, the Commission ultimately approved the addition of an essay at its fall 1985 meeting.

Reliability Issues Associated with Adding the Essay

Among the problems which contribute to the relatively low score reliability of direct measures of writing are: controlling topic variability, maintaining score scale stability, and maximizing reader consistency in applying a score scale. Research has suggested that to attain comparable score reliability for a direct measure of writing to that of a multiple-choice test, an essay test would require at least six essays, each scored by at least two readers (Breland, 1986). Early on, it was evident that, given the costs associated with scoring essays, only one essay could be given as part of the GED Tests. Consequently, it was clear that the essay would have to be incorporated with a multiple-choice section in order to obtain an acceptably reliable measure of writing skills. Given the decentralized nature of the GED Testing Program (many testing sites and testing dates) and the traditional values of local autonomy and rapid return of test results to examinees, it was also clear that many essay scoring sites would be required.

Integrating an Essay into the GED Writing Skills Test

Initially, the GED Testing Service: 1) reviewed possible multiple-choice writing skills item types and test formats, 2) reviewed research on essay scoring rubrics, 3) reviewed and researched types of essay topics and controlling variability due to topics (GEDTS, 1985a), and 4) researched comparisons of reading reliabilities across multiple scoring sessions (GEDTS, 1985b). In this presentation, the emphasis will be on establishing and maintaining scale stability and reading reliability; because it was necessary for GEDTS to combine the essay (direct assessment), with a multiple-choice section (indirect assessment), included in this paper is specific information related to that task.

Multiple-Choice Test Formats and Item Types

The review of writing skills multiple-choice test formats and item types resulted in choosing a different test format and different multiple-choice item types than are on the GED Writing Skills Tests used before 1988. A description of the new test format and item types along with examples is given in Appendix B (excerpted from GEDTS, 1985c). The new

test format and item types were chosen to more closely simulate the writing process and, hopefully, increase the correlation between multiple-choice and essay scores.

Phases in Developing and Maintaining Standard Scores

Because the nature of the GED Testing Program requires that essays be scored in decentralized scoring sessions, a major concern is the consistency with which essay readers apply the scoring guide standards. In research comparing reading reliability and score scale stability across scoring sites (GEDTS, 1985b), it became evident that, while respectable reading reliabilities and score scale stability could be attained, rigorous procedures for controlling and monitoring essay scoring at the various scoring sites would have to be applied. Among the procedures critical to obtaining high reading reliabilities and maintaining scale stability are: 1) training and certification of readers (see Appendix D for a discussion on reader training), 2) a certification process that evaluates whether a site is performing well enough to score essays operationally, and 3) scoring site monitoring procedures to evaluate whether the site's ongoing performance is sufficient to warrant continued certification to score operational papers. The GED Writing Committee contributes very carefully selected anchor papers and training papers to each of these three procedures, thus providing a "standard" against which readers and scoring sites can be evaluated.

The first and most essential task in developing and implementing a program that requires that essays be scored in hundreds of decentralized scoring sites and sessions, is the establishment of strong scoring standards. GEDTS has employed the following strategies to insure valid and reliable standards.

The GEDTS Writing Committee

The establishment of a permanent GEDTS Writing Committee was the most critical phase of the process. This Committee, composed of university, adult education, and high school instructors, provides "expert judgement" validity for the assessment. Because the Committee members are directly involved in writing instruction on a day-to-day basis, they are familiar with instructional expectations for candidates at particular performance levels. Committee members also have a historical perspective on the development and implementation of the essay component; providing continuity from administration to administration. This consistency in applying the scoring criteria to essays across administrations is essential to insuring that the papers are scored accurately over time.

In addition to sustaining scale stability, and aside from the development of the scoring guide, ongoing roles of the Committee include: 1) reviewing and evaluating potential topics, 2) selecting operational topic anchor and training papers which are used in training Chief readers and readers, 3) selecting operational topic anchor and site certification and

monitoring papers, and 4) selecting anchor and a minimum of 12 sets of recalibration papers for each operational topic. In order to select the necessary papers, periodic Writing Committee meetings are held. Members are first reoriented to the GED Official Scoring Guide through the scoring of a recalibration set to which they previously assigned scores. Committee members are then instructed to score essays based on the descriptions included in the scoring guide. This is necessary because even the Writing Committee, with 100 percent agreement, can drift from the scores as defined by the scoring guide. It is only by relating the selection of the score to the description of that score point in the scoring guide that this can effectively be detected. Other roles of the Writing Committee are referenced in the discussion below.

Choice of a Scoring Rubric and Development of the Official GEDTS Scoring Guide

Of the scoring rubrics investigated, holistic scoring was chosen as the method of evaluating GED examinee essays. An explanation of the rationale for choosing holistic scoring is given as Appendix C (excerpted from GEDTS, 1985a).

To insure greater clarity, consistency, and validity for essay scoring scale stability across topics, as well as within and across scoring sites and sessions, a generic scoring guide was developed. The essay component of the Writing Skills Test is norm-referenced--designed to provide a basis for comparing the performance of GED candidates with the performance of a national stratified random sample of graduating high school seniors. Because of this, the GED design of the essay scoring guide had to be descriptive, rather than prescriptive, in nature.

The Writing Committee was asked to rank order essays in six stacks based on abilities, and then to describe, in general terms, the characteristics of papers at different points on the scoring scale. Appendix D (excerpted from GEDTS, 1985b) further describes how the scoring guide was inductively developed from characteristics of high school senior essays written on one "anchor" topic. As Appendix D points out, the decision to apply one scoring guide across all future topics and readings helps to provide a foundation for scale stability across topics, readers, and scoring sessions. The scoring guide that was developed is given as Appendix E. The scoring standards as identified by the Committee are not prompt-specific, but reflect the original intent of the scoring guide and Committee based on the description of the actual graduating high school students' essays. These are invariant standards that remain in place over the years until a new norming is conducted.

Evaluating and Selecting Uniform Topics

To help minimize the problem of controlling topic variability, it was decided to restrict GED essay topics to one type (i.e., expository). An example of a topic is given in Appendix F (excerpted from GEDTS, 1985c). Judgmental criteria were developed to help the GED Writing Committee evaluate potential topics. Appendix G (excerpted from GEDTS,

1986) enumerates many of these judgmental criteria.

After potential topics are reviewed and edited by the Committee, they are field tested in administrations to both high school seniors and GED examinees. Papers written on the potential topics from these administrations are evaluated in a topic selection reading. The topic selection reading focuses on whether, in light of actual papers, the topics meet the criteria presented in Appendix G.

If topics "survive" these judgmental evaluations, statistical evaluations are performed. For high school seniors' scores on the set of papers, a Kruskal-Wallis test of the homogeneity of score distributions and a test of the equality of correlations between multiple-choice and essay scores are performed. To support the administration of any operational essay topic in combination with any multiple-choice test form, only those topics that yield non-significant results on both statistical tests are retained for operational use.

Selecting and Training Chief Readers and Readers

The nature of the GED Testing Program requires essays to be scored in decentralized scoring sites and sessions: it is imperative that all scoring sites have a GED trained and certified Chief Reader. Each jurisdiction's GED Administrator is responsible for selecting a potential Chief Reader based on GED requirements. GED Chief Reader qualifications and certification requirements are included in Appendix H.

The decision was made to bring all potential Chief Readers to a central location for training to help enforce the concept that all scoring sites are part of one program and must apply GED scoring standards uniformly. The training phase involves a two-and-a-half-day modified holistic training and qualifying session conducted by the GED Writing Assessment Specialist with some assistance from the GED Writing Committee. The major role of the Writing Assessment Specialist in training the Chief Readers, and the Chief Reader in training readers, is to communicate accurately the intent of the GED Official Scoring Guide and to focus discussion on attaining an understanding of the rationale of the Committee assigned scores as related to the Scoring Guide.

The training consists of an orientation to the background and purpose of the assessment, a thorough review of the training manual, discussion of reader objectivity issues (Appendix I), guidelines for holistic scoring (Appendix J), practice in scoring papers sets, and certification as Chief Readers.

The training papers--five sets of ten papers, and the certification papers--two sets of 20 papers, are standard papers as identified by the Writing Committee. This was accomplished by making enough copies of a high school student writing sample so that all the members could read and discuss the same paper at the same time. Committee members independently scored each paper; the score was then compared to other Committee member scores. A paper was assigned a Committee score and included as a

training paper only if at least 80% of the members assigned the same score. In addition, Committee comments about why each paper received the score it did are included to serve rationale statements during the training.

The training sets were then designed not as random paper sets with equal distribution at score points, but with a disproportionate number of borderline papers, high and low, at each line to train the readers to correctly discriminate between scoring points. Because the mean score on the GED essay for high school seniors is 3.5, and the mean score for GED candidates is 2.5, the training sets consisted of a larger number of essays assigned scores of 2, 3, and 4. In addition, the training sets included examples of the problematic papers that caused extensive deliberation and would be the probable cause of most third readings (i.e. rap songs, abstract ideas.)

With the initial training set, potential Chief Readers and readers are guided through a simulated Writing Committee essay selection exercise conducted to allow the participants to articulate score selection in the "language" of the scoring guide. This exercise serves two purposes. First, it shows the potential readers the procedure involved in selecting papers and the importance of using the scoring guide in the selection. Second, it allows the trainer to hear the rationale a potential reader is using when assigning a score point. This can be beneficial to a Chief Reader who must select an alternate Chief Reader.

To qualify as a reader and as a Chief Reader, each person must demonstrate the ability to apply the GED essay scoring standards by rating reader certification paper sets with a 50 percent exact agreement and 50 percent contiguous agreement with pre-assigned Committee scores.

The training session serves as a model for the Chief Readers to use when training potential readers for each scoring sites. The Chief Readers are instructed to use the same materials in the same order as presented in the Chief Reader training session. This is done to eliminate as many variables as possible in determining site certification. Essay reader qualifications and certification requirements are also listed in Appendix H.

Phase 5: Site Certification

After the Chief Reader and all readers for a particular scoring site have been certified, GED requires the site to be certified. Each scoring site must demonstrate the ability to score essays in accord with the standards defined by the GED Testing Service. To qualify for certification, a site must demonstrate the quality of its essay scoring in two respects:

- 1) reading reliability - the consistency with which different readers in a given reading award the same scores to a given set of essays, and
- 2) scale stability - the degree to which readers uniformly

apply the GEDTS score scale in evaluating essays both within and across scoring sessions. This will be determined by matching the total score (of two or, when necessary, three readers) awarded an essay by the potential site with the total score awarded an essay by GEDTS. This will be determined by matching the total score (of two or, when necessary, three readers) awarded an essay by the potential site with the total score awarded an essay by GED Writing Committee.

All essays included as site certification essays are original essays written by high school seniors or GED examinees. Only essays on which there was at least 80 percent agreement among the Writing Committee were included in the certification set. The majority of the papers had 100 percent agreement. There are thirteen sets of twenty papers; each set has an equal distribution among score points and agreement to eliminate any variables among readers.

Paper sets are assigned to reader numbers by the GED Writing Assessment Specialist; each reader must read two sets of twenty papers and each paper set must be read twice. If there are any discrepant scores, the Chief Reader must read the essays and assign a third score. Chief Readers do not know the GED Writing Committee scores for the site certification papers.

Site certification paper sets and scoring rosters are then returned to GEDTS for evaluation. Appendix J shows a sample table of actual results reported back to a scoring site after a site certification reading. After discussing the results and the implications of the results of the initial jurisdictions who completed site certification, GEDTS established the following site certification standards:

PERCENTAGE OF AGREEMENT WITH GEDTS TOTAL ESSAY SCORES*	STATUS	IMPLICATIONS
90-100%	Pass	The site meets the standards as established by GEDTS and may begin scoring essays.
85-90%	Provisional Pass	The site conditionally meets the standards as established by GEDTS and will be allowed to begin scoring essays with the understanding that additional training and a second certification must occur within 60 days

after receipt from GEDTS of a diagnostic report of the scoring site and readers.

84% and below

Failure

The site does not meet the standards as established by GEDTS and will not be allowed to score essays until remediation and a second certification reading is conducted and a passing score is obtained. GEDTS will send a diagnostic report identifying the problem areas and/or readers.

* AGREEMENT WITH GEDTS TOTAL ESSAY SCORES INCLUDES SCORES THAT ARE THE SAME AND SCORES THAT ARE CONTIGUOUS WITH THE GEDTS TOTAL SCORES.

The major concern of GED is the agreement of the reader with the GED Writing Committee assigned score. It is possible for some sites to have ninety to one hundred percent inter-rater reliability but only have sixty percent agreement with the GED assigned score. Because scoring is decentralized, it is critical to relate the reader and site scores with the GED Writing Committee standard scores to prevent a natural drift to local or individual scoring standards. A diagnostic report evaluating individual reader as well as site performance is returned to each Chief Reader.

Monitoring

The second major task in operating a program that allows decentralized scoring is to reduce or optimally eliminate scoring drift. Kadzin (1977) defined consensual observer drift as "the tendency of observers to change the manner in which they apply the definitions of behavior over time." Because this can impact scoring accuracy or scale stability without affecting inter-rater reliability, GED instituted three monitoring strategies to insure score stability and eliminate drift from scoring site to GED standards: Chief Reader monitoring, recalibration, and site monitoring.

Because GEDTS can monitor scoring sites on a periodic basis only, the Chief Reader's role as monitor is critical. The Chief Reader can check on each reader from session to session, thus discerning if readers consistently adhere to the GEDTS scoring standard. Since the Chief Reader resolves all discrepant scores, a record can be kept each time a reader assigns a score that the Chief Reader feels is off-scale. If any reader has more than one discrepant score during any scoring session,

the Chief reader must notify the reader of the scoring drift and help the reader get back on scale.

Scoring accuracy is also monitored on a scoring session basis by having readers score a set of five recalibration papers that were prescored by the Writing Committee. Readers are made aware of their scores and are cautioned to check that they remain consistent with the committee scores. Readers are also checked to determine if they are rating recalibration sets consistently too high or too low. Therefore, not only reader accuracy to the score point but reader drift in either a high or low direction is monitored by using the recalibration paper sets. The GED Writing Committee has developed a minimum of thirteen recalibration sets for each operational topic.

Finally, GEDTS will conduct site monitoring on a periodic basis by sending monitoring paper sets to the Chief Reader with instructions to include the monitoring exercise during the next scheduled scoring session. This monitoring will be essentially the same as the site certification procedure with the same guidelines and criteria for passing. Again, it is essential that the readers agree with the GED standard scores assigned by the Writing Committee. This will be a quality check to compare the scoring sites inter-rater reliability with the GED scoring standards. Only then can GED conclude that a scoring site has maintained scale stability and scoring accuracy.

Derivation of the Writing Skills Test Weighted Composite Standard Score

Procedures discussed above form the foundation of addressing score reliability of the reported weighted composite scores for the GED Writing Skills Test. That is, because reading reliability and score scale stability directly impact score reliability, they must be established before one can hope to attain adequate weighted composite score reliability.

The explication that follows, although not specific to this presentation, details the process needed to incorporate a direct measure of writing into the GED Writing Skills Test by drawing upon experience in assisting one jurisdiction that, by legislative mandate, had to accomplish the task prior to the 1988 introduction of an essay for all GED examinees. Because revised Writing Skills multiple-choice tests and normative scales were not yet available, the decision was made to incorporate an essay as part of the current Writing Skills Test. (The revised GED Tests for 1988 were standardized in an administration in U. S. and Canadian high schools.) Even so, procedures described below were used, for the most part, in forming the score scale which will combine the revised multiple-choice and the essay component for the 1988 GED Writing Skills Test.

In determining the relative weights for the essay and multiple-choice sections, several considerations were pertinent. One was that the GED Advisory Committee had stated that the essay should be weighted heavily;

specifically, the committee recommended a relative weight as high as 50 percent. However, the committee, informed of reliability concerns, indicated a willingness to accept a relative weight as low as 35 percent. After some discussion, it was determined that the GED Testing Service staff should determine a relative weight within this range that would weight the essay as much as possible without diminishing the estimated test-retest reliability below a level professionally acceptable for tests in the nature of the GED Tests. Operationally, the staff decided to establish relative weights for the essay and the current multiple-choice test such that the estimated weighted composite score reliability would be no less than the test-retest reliability for the current multiple-choice Writing Skills Test taken alone. (Typical Writing Skills Test-retest reliabilities are 0.90 for seniors and 0.85 for GED examinees.)

Topic Selection

As part of a 1985 equating administration to a national stratified random sample of graduating high school seniors, four new forms of the current GED Tests and the original anchor form of the pre-1988 version were administered. Fifteen potential essay topics (lettered A through O) that had been reviewed by the Writing Committee were paired with the total of five Writing Skills Test forms built to the then current specifications (not the specifications for the 1988 GED Tests). Of these 15 topics, one was the anchor topic described earlier in connection with the development of the scoring guide. Because the anchor topic had been a borrowed "tried and true" topic from another testing program and could not be considered secure, it was not eligible for operational use with GED examinees. Of the remaining 14 topics, six were rejected (or later revised and retried) during the topic selection reading process (described earlier). The remaining seven topics (B, C, F, G, I, J, and N) were then subjected to a Kruskal-Wallis test for homogeneity of score distributions and a test of their correlations with the multiple-choice measure. Topic G was rejected because both its score distribution and correlation to the multiple-choice measure were found to differ from those of the other topics. Topic I was rejected on the basis of the Kruskal-Wallis test alone. Pages one and two of Appendix L show the comparisons of correlations and distributions.

The standardization administration for the series of GED Tests to be used by this jurisdiction was conducted in 1980. The presence of the anchor form in both the 1980 standardization and the 1985 equating administrations provided for the scaling of the new forms to the 1980 normative scale. For the project to incorporate an essay into the current Writing Skills Test, essay raw scores were correlated with the scaled multiple-choice standard scores. A test for differences in ability (as estimated by multiple-choice standard scores) across groups which took a given multiple-choice and essay combination was done, and no differences were found (see page 3 of Appendix L). Therefore, it was decided that because the five multiple-choice forms scores were all on an equated (equipercentile) standard score scale, the data for topics paired with multiple-choice forms would be pooled across multiple-choice forms

for the correlational analyses. Further, it was decided to convert the essay raw scores to the same standard score scale as that for the 1980 graduating seniors.

Operational Topic Equating

With the pool of eligible topics narrowed to five (B, C, F, J, and N) equipercentile and linear equating procedures were applied to the topic raw scores. Even though these topics had been found to have non-significant score distributions by the Kruskal-Wallis test, this was done to scale out very minor differences in score distributions found in the sample data. Plotted on normal probability paper the distribution of two of the topics (B and C) were nearly identical. To take advantage of a pooled sample size of 1,231 (620 on topic B and 611 on topic C) on these two topics, the remaining three topics were scaled to a combined B and C curve. (Appendix M contains the equipercentile plots of topics.) Because the equating relationships were judged to be sufficiently linear (especially in the region of the score scale where passing and failing decisions are made; i.e., about the 30th percentile), the linear equating results were used.

Solving for the relative weights

Adapting the formula from Nunnally (1978, p. 250), and plugging in a previously estimated 0.868 test-retest reliability for the multiple-choice test, a 0.60 test-retest reliability for the essays, and a correlation of 0.586 for these essays and the pooled multiple-choice forms standard scores, provided estimates of the weighted composite reliabilities given various relative weights in the range acceptable to the GED Advisory Committee. Table 1 below shows the estimated reliabilities given various essay weights.

Table 1

essay weight	estimated composite score reliability
.3	.878
.35	.871
.36	.869
.37	.867
.4	.861
.5	.832

Splitting hairs, a relative weight of 0.36 for the essay was chosen because it fit the arbitrary judgment not to weight the essay more than would diminish the estimated composite reliability below that of the multiple-choice test alone (0.868).

Given the relative weight of 0.36 for the essay, it was appropriate to

solve for the operational weights that would maintain the variance of the composite standard score scale at the desired standard deviation of 10. Using only the relative weights would not have maintained the desired variance. Adapting the formula for the variance of a weighted composite and given that both the multiple-choice and essay scores were scaled to be normal, mean 50, standard deviation 10, allowed for solving the equations:

$$\frac{W_{es}}{W_{es} + W_{mc}} = 0.36 \quad \text{and} \quad 1 = W_{es}^2 + W_{mc}^2 + 2(r_{mc,es})W_{es}W_{mc}$$

where: W_{es} is the essay weight,
 W_{mc} is the multiple-choice weight, and
 $r_{mc,es}$ is the correlation of the essay with the multiple-choice

Solving these equations yielded operational weights of 0.4001 for the essay and 0.7114 for the multiple choice standard scores.

Given the weights, the multiple-choice standard score, and raw essay scores, one can then employ a formula of the form shown in Appendix K for calculating the weighted composite standard score for the combined current GED Writing Skills Test and an essay written on one of the five operational essay topics

Discussion and Conclusions

When direct assessment of writing is used, it is mandatory to establish and implement scoring standards and to initiate monitoring procedures which provide continuity among scoring sites and scoring sessions. The scoring process for all GED scoring sites is continually evaluated: both on a regular basis by the Chief Reader and systematically by the GED Testing Service. Insuring a fair and stable scoring from administration to administration requires careful planning and systematic procedures. These conclusions are applicable to both small programs and much more complex programs such as the decentralized GED Testing Service situation.

Will the process described herein result in acceptably reliable GED Writing Skills scores? To help answer this question, some examinees are being asked to take two forms of the new GED Writing Skills Test. From this data, empirical test-retest reliabilities will be calculated.

One judgment from the experience gained so far is clear: constant vigilance is required to maximize essay score scale stability and reading reliability. This vigilance is essential given that achieving acceptable test-retest score reliability depends, in large measure, on first establishing essay scoring scale stability and reading reliability.

References

- Breland, H. M. (1986, April). Study objectives and outcomes of reliability and predictive validity analyses. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- General Educational Development Testing Service. (1984). GED tests specifications committee report. Washington, DC: American Council on Education.
- General Educational Development Testing Service. (1985a). The relationship between scores on the GED writing skills test and on direct measures of writing (Research Studies No. 6). Washington, DC: American Council on Education.
- General Educational Development Testing Service. (1985b). Adding an essay to the GED writing skills test: reliability and validity issues (Research Studies No. 7). Washington, DC: American Council on Education.
- General Educational Development Testing Service. (1985c). The 1988 tests of general educational development: A preview. Washington, DC: American Council on Education.
- General Educational Development Testing Service. (1986, January). GED essay topics: Some specifications . . . and an invitation. GED Items. p. 3.
- Nunnally, J. C. (1978). Psychometric Theory (2nd ed., p. 250). New York: McGraw-Hill.
- Spille, H. A. (1984). Summary of surveys concerning future GED tests (phase 2). (General Educational Development Testing Service Memorandum No. 31). Washington, DC: American Council on Education.

The Tests of General Educational Development

The GED Program

The General Educational Development (GED) Program began in 1942 with the development of a battery of tests for U.S. military personnel who had not completed their high school studies. The Tests of General Educational Development (GED Tests) provided an opportunity for these people to demonstrate that they had acquired the learning outcomes expected of a high school graduate. Because of this program, many people were able to qualify for jobs and pursue postsecondary education upon leaving the military service.

After the war, it became apparent that civilians could also benefit from this program, a need the American Council on Education (ACE) undertook to fill. Since that time, ACE has directed the GED Testing Program. In the 1950s, several states began to recognize successful GED candidates by issuing credentials based on the examinations. These credentials were considered equivalent--for purposes of occupation, military service, and college admission--to a high school diploma earned at the end of formal study.

From this modest beginning, the GED Testing Program has grown considerably. Today, the tests are used in all 50 states, U.S. territories, and in most Canadian provinces. In 1984, more than 700,000 people took the tests, and about 73 percent qualified for a credential according to standards set by their states, territories, or provinces. Over the years, the GED Tests have provided personal satisfaction as well as occupational and academic opportunities for millions of adults who, for many reasons, were unable to complete their formal high school studies.

The GED Tests' Content*

The GED Test battery includes five tests, each measuring the skills considered to be the major outcomes of a high school education. Because GED examinees typically range in age from 16 to 90 and have acquired their skills in a less formal manner than the high school student, the tests focus on the use of major skills and concepts rather than on recall of specific facts. The test questions focus on the general abilities to analyze, evaluate, and draw conclusions. A brief description of the tests follows.

Test 1. Writing Skills--Part I contains multiple choice questions intended to measure a person's ability to use Standard Written English clearly and effectively. The examinee must identify and correct errors in spelling, capitalization, punctuation, diction and style, and usage, as well as recognize the preferred ways of writing sentences. Part II provides an expository essay topic about which examinees are expected to have general knowledge. Topics generally require examinees to present an opinion and defend it with appropriate evidence, detail, and arguments.

Test 2. Social Studies--Material is drawn from United States and world history, economics, geography, political science, and behavioral sciences. The questions require the candidate to demonstrate an understanding of the primary concepts and principles in these areas by understanding the meaning of information (written and graphic), applying information and ideas, distinguishing fact from opinion, drawing conclusions, identifying causes and effects, and making judgments.

Appendix A

Test 3. Science--Material is drawn from the areas of biology, earth science, chemistry, and physics. The questions require examinees to demonstrate an understanding of the primary concepts and principles in these areas. Like the Social Studies test, most questions require using information presented to analyze and solve problems, explain results, interpret findings, and/or predict outcomes.

Test 4. Interpreting Literature and the Arts--Excerpts are drawn from popular or classical literature (prose non-fiction, fiction, poetry and lyrics, or drama) or commentary about literature and the arts. The questions require examinees to demonstrate an understanding of what is read, interpret the meaning and intent, and/or draw conclusions. Typically, questions require examinees to go beyond what is directly stated and draw inferences from what is implied.

Test 5. Mathematics--Questions cover the areas of arithmetic (measurement, number relationships, and data analysis), algebra, and geometry. Information is presented in graphs, charts, tables or diagrams which generally depict a realistic situation. Typically, questions require some computation to solve the problem or draw conclusions.

The five tests are developed by committees of educators in each subject area under the direction of the GED Testing Service staff.

Keeping the GED Tests Up to Date

The series of tests currently in use was standardized on a sample of high school seniors in 1977 and first introduced in GED Testing Centers in 1978. A second standardization administration in high schools was conducted in 1980. New forms of the tests within this series were

developed annually, and score scales were established on the basis of high school senior performance.

A revised series of tests is now being developed. The major revisions include an essay as part of the Writing Skills test and increased emphasis on higher level thinking skills on all five tests. The first of this revised series of tests is being administered in high schools in spring 1986 and will be introduced in GED Testing Centers in 1988.

The periodic administration of GED Tests to high school seniors provides the basis for comparing the performance of GED examinees to that of high school seniors and establishes the basis for evaluating examinees' scores for the purpose of awarding credentials.

For Further Information

For further information about the GED Tests or Testing Program, please write to your local GED Testing Center, state department of education, or to the GED Testing Service, American Council on Education, One Dupont Circle, NW, Washington, DC 20036-1163.

*The GED Tests described here are to be introduced in GED Centers in 1988. See discussion under "Keeping the GED Tests Up to Date."

Writing Skills, Part I

The Writing Skills test of the revised GED Tests is comprised of two parts, one part to test knowledge of the conventions of written English, and the second to test directly the ability to write. The scores earned on both components of the test will be combined and reported on a standard score scale as a single score. The recommendation that the test include a writing sample is a dramatic revision to the current test content. However, the Language Arts panel (and, indeed, the entire test specifications committee) expressed its view on this matter in no uncertain terms: "The English specifications panel believes that no one should receive credit for high school equivalency without being asked to demonstrate writing ability directly as well as indirectly."

Content

Part I of the Writing Skills test will resemble the current GED Writing Skills test in several ways. First, both the current test and a part of the revised test use a multiple choice format which presents items that require error recognition and correction. The revised test will include items in approximately the following proportions:

- Sentence Structure: 35%
- Usage: 35%
- Mechanics (spelling, punctuation, and capitalization): 30%

In the current Writing Skills test, 30% of the items are Sentence Structure/Diction and Style, 30% are Usage, and 25% are Mechanics. The remaining 15% in the current test consists of items in Logic and Organization, which will be measured directly in the essay component of the revised Test One.

Context

The subject matter for the Writing Skills test, Part I includes those topics with which the examinee is likely to be familiar. Relevant subjects may include topics of practical interest such as consumer affairs, career skills, and computer awareness or subjects of more general interest such as vacations, health, family life, and education.

Format

As adults usually need to write and proofread connected pieces of discourse, Part I directly measures proofreading skills in the more realistic setting of paragraphs. These paragraphs are 10-12 sentences long and, when corrected, are examples of good writing.

Some of the sentences within each paragraph contain errors which students must locate and correct. In some cases the sentences do not contain errors but require revision for the sake of clarity or logic. The errors to be corrected are

those most often encountered in student writing, most complained about by the public, and most likely to hamper a person's advancement, acceptance, or ability to communicate effectively to various audiences. The sentence to be corrected is repeated in each item and the alternatives (or answers) are presented in the order in which they occur in the sentence.

In at least half of the items (or questions), the alternatives do not appear by categories. That is, alternatives for each item may often be from any of the five content areas—again in an attempt to create a realistic proofreading situation. (This range of alternatives is found in the sentence correction item type described below.)

Item Types

The items in the Writing Skills test, Part I are classified by item type. "Item type" refers to the *method* by which the content areas are tested and by which the examinee is asked to respond. The three item types are used in the following proportions:

- (1) Sentence Correction: 50%
- (2) Sentence Revision: 35%
- (3) Construction Shift: 15%

Examples of each type are provided below.

Item Type 1: Sentence Correction

This item type is used to test skills in all content areas of the Writing Skills Test, Part I. The sentence correction item type repeats one of the sentences from the paragraph and asks what correction should be made. Several examples are given of this type with different content areas and categories.

Item Type 2: Sentence Revision

This item type is used to test skills in the content areas of Sentence Structure, Usage, and Punctuation. The sentence revision format consists of a stem with a part underlined that may or may not contain an error in sentence structure. The alternatives which follow list various possible corrections to the underlined section of the stem. In this item type, the first alternative is always exactly the same as the original sentence.

Item Type 3: Construction Shift

This item type is used to test skills in the content areas of Sentence Structure and Punctuation. The construction shift type consists of a stem which must be rewritten using a different structure. The sentence which results from the revision must have the same meaning as the original sentence and must be correct and clearly stated. While the stems in sentence correction and sentence revision item

types usually contain errors, construction shift items do not. These items aim at testing an examinee's ability to employ alternate structures correctly rather than the ability to correct errors. Implicitly, logic skills are also tested here, as well as an understanding of sequence of events.

Cognitive Levels

All items tested in the Writing Skills test, Part I are at the application level and classified by item types, as described above.

The following is a sample paragraph appropriate for Part I of the Writing Skills test by virtue of its subject, length, and reading level. Following the paragraph are sample items illustrating a variety of item types, content areas and categories.

SAMPLE PARAGRAPH

Directions: The following items are based on a paragraph which contains numbered sentences. Some of the sentences may contain errors in sentence structure, usage, or mechanics. A few sentences, however, may be correct as written. Read the paragraph and then answer the items based on it. For each item, choose the answer that would result in the most effective writing of the sentence or sentences. The best answer must be consistent with the meaning and tone of the rest of the paragraph.

(1) Computer programs are sets of specialized instructions. (2) Programs, stored on magnetic disks, are also known as software, the hardware is the computer itself. (3) A program talks to a computer. (4) It tells a computer how to perform certain tasks, such as word processing or accounting. (5) The computer reads the program disk electronically by spinning it. (6) When you type on the keyboard you provide the computer with information and commands. (7) The program manual explains which keys, or combinations of keys, will activate these commands. (8) For instance, in word processing there are commands to delete words and sentences. (9) The program translates what you type into electronic impulses that the computer understands. (10) At the same time, the program translated electronic impulses from the computer into words and numbers on the screen. (11) With this technology, businesses can rely on computers for everything from inventories to accounting. (12) At home, individuals can write letters balance the family budget, and even shop for groceries. (13) The capabilities of computers are changing every day and with them the way we live our lives.

Sample Test Items

Content: Sentence structure
Category: Coordination/Subordination
Item Type: Construction Shift
Answer: 2

Sentences 3 & 4: A program talks to a computer. It tells a computer how to perform certain tasks, such as word processing or accounting.

The most effective combination of sentences 3 and 4 would include which of the following groups of words?

- (1) It talks to a computer
- (2) A program tells a computer how
- (3) and it tells a computer
- (4) or it tells a computer
- (5) A program performs certain tasks

Content: Sentence Structure
Category: Sentence Fragment
Item Type: Sentence Revision
Answer: 1

Sentence 5: The computer reads the program disk electronically by spinning it.

Which of the following is the best way to write the underlined portion of this sentence? If you think the original is the best way, choose option (1).

- (1) electronically by spinning it.
- (2) electronically. By spinning it.
- (3) electronically, by spinning it.
- (4) electronically; by spinning it.
- (5) electronically, however by spinning it.

Content: Usage
Category: Verb Tense
Item Type: Sentence Correction
Answer: 3

Sentence 10: At the same time, the program translated electronic impulses from the computer into words and numbers on the screen.

What correction should be made to this sentence?

- (1) replace *At the same time* with *Because*
- (2) remove the comma after *time*
- (3) change *translated* to *translates*
- (4) insert a comma after *impulses*
- (5) no correction is necessary

Scoring Method

Considerable attention was given to selecting the best possible method for scoring the writing samples obtained from the high school students. Holistic scoring was selected as the optimal method for this study for several reasons. First, holistic scoring is a quick and relatively simple procedure, so it is an efficient method for scoring large numbers of papers. Because the GED Testing Service wished to consider the feasibility of essay scoring as one of the research questions, holistic scoring was a technique preferable to either analytic or primary trait scoring, both of which demand of essay readers more training and time per essay. Second, to answer the primary research question involving the correlation between essay and multiple choice scores, high degrees of inter-reader reliability were required, and holistic scoring offered the most accessible method of attaining that reliability (Cooper, 1977). An additional factor in the scoring method decision was the acknowledgement that, if a writing exercise were to be added to the GED Writing Skills test, practical constraints would limit the exercise to a single topic. This limitation further suggested the use of holistic scoring and the combination of a holistic essay score with multiple choice scores to yield a composite writing score. The relative unreliability of a holistic score on a single topic (Charney, 1984; Cooper, 1977) is mitigated appreciably when the score is combined with a more reliable indirect measure of writing (Godshalk, et al., 1966).

Holistic scoring is a method of evaluating a piece of writing based upon a paper's overall effectiveness. Often referred to as "general impression scoring," the method relies upon the idea that the effectiveness of a piece of writing does not necessarily equal the sum of its parts, and that informed essay readers can come to agreement about the worth of a whole piece of discourse while they may disagree about the relative value of its component parts (Conlan, 1976).

Because holistic scoring is essentially a process of rank ordering papers (Cooper, 1977), it is frequently observed that one of the limitations of the scoring process is that the ranking indicated by the scores varies according to the population which comprises the sample. That is, a high school student's paper scored as a "six" in one reading may become a "four" if placed among papers written by college graduates because the paper's rank changed in relation to the universe of papers of which it was a member. To control this drifting of standards, a scoring guide was developed for each topic. Three separate guides were written to reflect the clearly different purposes suggested by each of the topics. Because the guides were created inductively—by describing the common characteristics of the agreed-upon model papers—they could be regarded as descriptive rather than prescriptive.

HOW THE GED ESSAY IS SCORED

GED Testing Service

Holistic Scoring

Each paper written for the GED Writing Skills Test is scored by two trained readers. The readers score the papers "holistically;" that is, the piece of writing is evaluated on the basis of its overall effectiveness. The salient features of holistic scoring are well described by Conlan (1976):

The basic assumptions of the holistic reading are that each of the factors involved in writing skill is related to all the others and that no one factor can be separated from the others. Readers must judge each essay as a whole; they must read each paper for the impression its totality makes. A misspelled word, a comma splice, a sentence fragment, a misplaced modifier should carry no great weight in scoring a paper. The candidate is entitled to make some mistakes: he or she is writing hurriedly in a tense situation, without recourse to a dictionary, without the customary time for deliberation. If readers read each paper as a whole, then, they are better able to judge the competence of the writer. If the paper is poorly written, that will be part of the reader's first impression; there is no need to analyze it word by word to decide that it is badly done. If the paper contains mistakes because the writer, though capable, has been forced to hurry, the reader will judge the general quality of the paper most effectively by reading it as a whole. The motto of the reading might well be "Read quickly and judge. Do not re-read."

The Six Point Scale

The scale used by each reader in the holistic evaluations ranges from one (low) to six (high). The two readers' scores are then added, resulting in a range of scores from two to twelve. If the two readers' scores differ by more than one point, the paper is scored a third time. The total score for papers scored by three readers is twice the average of the three scores.

Because the six point scale is an even-numbered scale, there is no midpoint. The lack of a midpoint forces readers away from a natural tendency to drift towards the middle. With an even-numbered scale, readers must decide whether each paper belongs in the upper half or the lower half of the papers being scored. Readers can make this decision only if they have some awareness of the total range of writing ability represented by the papers; this is provided by sample papers in conjunction with a descriptive scoring guide.

How the Standards are Defined

The standards for an essay scoring session are defined by the GED essay scoring guide and by sample papers illustrating the different points on the scoring scale. These standards were originally established by examining papers written by a national sample of high school seniors and will be maintained by following the same process for each topic that is used in the GED Writing Skills Test. A group of highly experienced readers reads papers in the pool of high school essays to establish the range of abilities demonstrated by the norming population of (high school) students. After they have selected papers which represent the full range of ability, the readers assign the top score to the best of the papers, the lowest score to the weakest papers, and corresponding scores to other papers throughout the range. The readers' judgments during this process are made independently, and the papers that ultimately serve as samples include only those on which the readers' scores agree. These sample papers represent the full range of abilities demonstrated by a national sample of high school seniors and thus serve as standard setters, or "rangefinders."

The GED Essay Scoring Guide

The GED essay scoring guide is a further articulation of these standards and accompanies the rangefinders in training sessions. Because it was developed inductively, the scoring guide is descriptive rather than prescriptive. That is, after the readers selected the rangefinders, they attempted to describe, in general terms, the characteristics of papers at different points on the scoring scale. While the sample papers illustrate standards only for the specific topic on which they are written, the scoring guide defines the characteristics papers should exhibit regardless of the topic the paper is written on. Although there is ample research demonstrating that different topics exhibit different strengths and weaknesses of a writer, a generic guide can be used if topics are developed with the scoring guide in mind. All topics developed for use in the GED Essay test are of the same rhetorical type and are of approximately equal length, reading level, and format. In addition, all potential topics are field tested and scored by trained readers, using the GED scoring guide to define standards. If readers find that papers written on a potential topic cannot be scored using the scoring guide, or that papers consistently

exhibit traits other than those described in the guide, the topic is rejected. A further check on the match of topic to scoring guide is made through a statistical analysis wherein the mean and variance of examinee scores on potential topics must approximate the mean and variance of the anchor topic for which the scoring guide was written.

Training of Readers

While the specific procedures for training readers and conducting the essay scoring session depend upon the number of papers to be scored and the number of readers participating, many of the basic principles remain the same regardless of the size of the effort. Readers are usually trained prior to or at the beginning of the essay scoring session. After an introduction to the principles of holistic scoring provided by a chief reader, readers are given the topic, the scoring guide, and the rangefinders, which include at least one paper at each point on the score scale. After a discussion of the tasks required by the topic, and a review of the qualities enumerated on the scoring guide, readers are asked to read the entire set of rangefinders, evaluating them swiftly on the basis of an overall impression and ranking them from best to poorest. Because the set of rangefinders includes more papers than there are points on the score scale (i.e., 8 papers for a six point scale), readers are instructed that they must use the entire range of scores. This is an important instruction in the training because many readers are reluctant to award scores at the top or at the bottom of the score scale. If readers are directed to give the top score to the best paper in the set of rangefinders, they are more likely to acknowledge that a "6" paper does not—and need not—represent perfection.

In a large group training session, the chief reader calls for a show of hands to indicate how readers scored the rangefinders. Ideally, all of the readers will have scored each of the rangefinders identically. In some readings, involving highly experienced readers, this ideal is realized. However, with inexperienced readers there are often substantial differences among readers' scores at this point. To help resolve those differences, readers designated as "table leaders" conduct discussions among the four to six readers at their tables to attempt to bring each individual to a point of consensus with the group. At the heart of the holistic scoring process is the necessity for each reader to be willing to accept—and in some cases consciously shift to—the standards defined in the scoring guide and adopted by the group. Following the discussion of the rangefinders, additional work is done towards this goal through the scoring of more sample papers, distributed singly or in sets of two or three. The scoring and

discussion of sample papers continues until the entire group of readers begins to show a consensus in their scoring. At this point, the training period ends and the actual scoring of essays begins, although table leaders may continue to work with individual readers if necessary.

Reliability and Stability

The goals of an essay scoring session are inter-rater reliability and reading "stability." Inter-rater reliability is the degree to which readers agree with each other; generally, the fewer papers which require a third reading, the greater the inter-rater reliability, though this index is affected by other variables as well. Reading stability is the degree to which papers are scored according to the fixed standards described in the scoring guide. This is a particularly important feature of an essay scoring session because the standards for scoring GED essays must remain fixed, regardless of when the essay is administered, where it is scored, or what specific procedures were used in the scoring session itself. It is important to note that a high degree of inter-rater reliability does not ensure reading stability. In short, just because readers are agreeing with each other on essay scores does not mean that they are assigning the scores according to the standards defined on the scoring guide.

To achieve these two goals, the reinforcement of scoring standards continues well after the initial training session is over. As the readers score papers, a table leader (in a small reading) selects scored papers at random to verify that the scoring is consistent with the definitions in the scoring guide. In cases of a disagreement, the chief reader or table leader discusses the paper with the reader. The monitoring process continues throughout the entire scoring session, and, in large readings, the chief reader periodically reviews the scoring of the table leaders.

When scoring large numbers of papers, readers generally read for periods from 30 to 45 minutes at a time and then take breaks. At the beginning of each reading period, additional sample papers are distributed and scored as an additional check on reader accuracy. Generally, readers score between 30 and 50 papers per hour, depending on various factors such as the reader's experience, the length of the papers, and the time of day. Readers are encouraged to read swiftly because a slow, deliberate reading of a paper tends towards analytical, rather than holistic, evaluation of writing. But readers are frequently reminded that accuracy is far more important than speed. Through this system of checks and re-checks, some assurance is gained that all readers are scoring according to the standards defined by the scoring guide and not according to one defined by each individual.

Papers will show some or all of the following characteristics.

Upper-half papers make clear a definite purpose, pursued with varying degrees of effectiveness. They also have a structure that shows evidence of some deliberate planning. The writer's control of English usage ranges from fairly reliable at 4 to confident and accomplished at 6.

- 6 Papers scored as a 6 tend to offer sophisticated ideas within an organizational framework that is clear and appropriate for the topic. The supporting statements are particularly effective because of their substance, specificity, or illustrative quality. The writing is vivid and precise, though it may contain an occasional flaw.
- 5 Papers scored as a 5 are clearly organized with effective support for each of the writer's major points. The writing offers substantive ideas, though the paper may lack the flair or grace of a 6 paper. The surface features are consistently under control, despite an occasional lapse in usage.
- 4 Papers scored as a 4 show evidence of the writer's organizational plan. Support, though sufficient, tends to be less extensive or convincing than that found in papers scored as a 5 or 6. The writer generally observes the conventions of accepted English usage. Some errors are usually present, but they are not severe enough to interfere significantly with the writer's main purpose.

Lower-half papers either fail to convey a purpose sufficiently or lack one entirely. Consequently, their structure ranges from rudimentary at 3, to random at 2, to absent at 1. Control of the conventions of English usage tends to follow this same gradient.

- 3 Papers scored as a 3 usually show some evidence of planning or development. However, the organization is often limited to a simple listing or haphazard recitation of ideas about the topic, leaving an impression of insufficiency. The 3 papers often demonstrate repeated weaknesses in accepted English usage and are generally ineffective in accomplishing the writer's purpose.
 - 2 Papers scored as a 2 are characterized by a marked lack of development or inadequate support for ideas. The level of thought apparent in the writing is frequently unsophisticated or superficial, often marked by a listing of unsupported generalizations. Instead of suggesting a clear purpose, these papers often present conflicting purposes. Errors in accepted English usage may seriously interfere with the overall effectiveness of these papers.
 - 1 Papers scored as a 1 leave the impression that the writer has not only not accomplished a purpose, but has not made any purpose apparent. The dominant feature of these papers is the lack of control. The writer stumbles both in conveying a clear plan for the paper and in expressing ideas according to the conventions of accepted English usage.
- * An asterisk code is reserved for papers which are blank, illegible, or written on a topic other than the one assigned. Because these papers cannot be scored, a Writing Skills Test composite score cannot be reported.

SAMPLE GED ESSAY TEST

Tests of General Educational Development Writing Skills, Part Two

Instructions

This is a test to find out how well you write. The test has one question which asks you to present an opinion on an issue or to explain something. In preparing your answer for this question, you should take the following steps:

1. Read all of the information accompanying the question.
2. Plan your answer carefully before you write.
3. Use the blank pages of this test booklet (or scratch paper) to make any notes.
4. Write your answer on the separate answer sheet.
5. Read carefully what you have written and make any changes that will improve your writing.
6. Check your paragraphing, sentence structure, spelling, punctuation, capitalization, and usage, and make any necessary corrections.

You will have 45 minutes to write on the question you are assigned. Write legibly and use a ballpoint pen so that the evaluators will be able to read your writing. The notes you make on the blank pages (scratch paper) will not be scored.

Your composition will be scored by at least two trained evaluators who will judge the paper according to its overall effectiveness. They will be concerned with how clearly you make the main point of your composition, how thoroughly you support your ideas, and how clear and correct your writing is throughout the composition. You will receive no credit for writing on a question other than the one assigned.

SAMPLE TOPIC

The automobile has certainly been responsible for many changes in the United States. Some of these changes have improved our lives and some have made life more difficult or unpleasant.

Write a composition of about 200 words describing the effect of the automobile on modern life. You may describe the positive effects, the negative effects, or both. Be specific, and use examples to support your view.

Refer to the steps on the directions page for more specific instructions about writing your composition.

All topics used with GED examinees during the studies remain secure test questions (for possible use in 1988) which cannot yet be published, nor can the examinees' responses. However, we are able to provide the example topic above, which was used with a national sample of high school seniors. The responses printed below were written by high school seniors; the scoring scale used to score the papers is the same as that used to score GED examinees' papers.

GED Essay Topics: Some Specifications...and an Invitation

Essay topics, like all other items to be used on the GED Tests, undergo a rigorous scrutiny before they are included on a final form. The process begins, logically enough, with the writing of "raw" topics that address the test specifications. In the case of the essay topics to be used on the GED Writing Skills test, the following general specifications must be met:

- The topic must be based upon information or a situation that is general enough to be familiar to most examinees. For example, a topic calling for a writer's views regarding the effect of automobiles on American life might well be appropriate, while one calling for an opinion of the space shuttle's value to the American economy would fail to meet this specification because it would require specialized knowledge. In short, the topic must offer access to as many examinees as possible—ideally, to all.
- The topic must offer an idea that examinees view as worth writing about. In this sense, the situation defined by the topic's context should be realistic. It would be a simple matter to construct contexts consisting of hypothetical issues, but these could not be expected to fully engage the writers. The purpose of the topic is to elicit a sample of writing which displays the examinee's strengths; the kind of topic most likely to do this is one which presents a situation about which the examinee already has an opinion or some feelings, or can form them readily.
- The topic should not elicit an overly-emotional response from the writer. While the situation provided as a prompt should trigger the writer's interest in some way, it should not go too far. The larger the population writing on a topic, the greater the chance that the topic will elicit an emotional response from someone. Emotional writing is rarely controlled writing and thus does not present an example of an examinee's best writing skill. For this reason, emotionally-charged issues, such as abortion, are not used as writing prompts.
- The topic should be clearly stated and contain only the amount of

information necessary to provide the prompt for writing. When examinees finish reading the topic, they should know exactly what they are being asked to write about. The objective is to state the topic clearly enough that students immediately know *what* they are going to write and spend their time working out *how* they are going to write it.

Only topics which meet these specifications become eligible for field-testing with GED examinees and with high school seniors. After topics are field-tested, the papers written on topics are holistically scored by experienced readers who use the GED Essay Scoring Guide (see *GED Items*, II.4). This type of "scoring session" differs significantly from one concerned solely with producing scores for essays. In this "topic selection reading," readers are asked not only to score the papers using the Scoring Guide, but also to evaluate how well the topic is working. Readers evaluate topics according to the following specifications:

- The topic should elicit papers with characteristics comparable to those described in the essay scoring guide. If, for example, the best papers written on a field-test topic seem significantly less accomplished than those described in the six category in the scoring guide, the topic may be more difficult for writers than desirable and therefore unusable. In the interests of making all of the topics as equal as possible in difficulty, all papers produced by the topics must conform to the scoring guide. If the papers exhibit characteristics that are clearly different from those described in the scoring guide, the topic will be rejected; the scoring guide will not be revised to conform to a new topic.
- The topic should elicit papers which illustrate the full range of student writing ability. A good topic allows strong writers to display their skill in writing, yet still allows weak writers access to the question. Because the field tests involve significant numbers of examinees and seniors, the total pool of papers produced for a topic should include essays at all points on the six point scale.
- The topic should elicit papers which clearly address the question provided. If papers written on a topic consistently show that writers are unclear about the question

writers are writing on a topic other than the one asked, the topic is ineffective. A topic which fails to meet this criterion was probably not clearly stated to begin with.

- The topic should elicit a variety of responses. To say that all papers must address the topic is not to say that all papers must appear the same. A good topic will yield papers whose ideas mirror the diversity of the population tested. Papers with a wide range of ideas will be less tedious for readers to score, and thus readers are likely to score more accurately. While scoring large numbers of papers inevitably becomes tiresome, a good topic produces papers that engage readers' interest to the extent possible.
- The topic should elicit fully-developed responses. A topic which yields a large proportion of incomplete papers may be too demanding for the time allowed. If papers consistently exhibit a shallowness of thought and inadequate development, examinees might not have enough information immediately at hand to write well about the topic.
- The topic should not elicit an emotional or biased response from readers. In the same way that care is taken not to trigger emotional reactions from writers, the resulting papers must not fuel preconceptions or biases among readers. Essay readers are urged to evaluate the writing, not the writer or the writer's values, and the scoring process contains numerous checks of this standard. However, readers cannot be expected to remain immune from emotional reactions, so an effective topic produces few papers that are likely to set off an emotional response in readers.
- The topic elicits papers for which readers can readily agree on scores. Readers should be able to distinguish upper half (6,5,4) from lower half (3,2,1) papers with relative ease. Where these distinctions are blurred, the topic itself may well be at fault. Among the many statistical checks on a topic's performance is the rate at which readers agree on scores for particular papers; a significant rate of disagreement among readers often indicates that the topic is yielding papers that cannot reliably be scored using the standards provided.

(continued)

Topics

At this point, you may be wondering how we ever hope to produce even a single operational topic that passes all of these stipulations (along with a few others relating to topic equating). In the spring of 1985, GEDTS field tested fifteen topics with a national sample of high school seniors, and six of the topics were judged acceptable for use in operational forms. To produce the fifteen for field testing, however, we began with well over a hundred topic ideas.

Revised
April, 1987

ESSAY READER QUALIFICATIONS:

To apply for certification as a GED essay reader, a candidate must possess the following:

- * a baccalaureate degree;
- * at least two years' total experience teaching English-language arts at the secondary, adult, or post-secondary level;
- * the ability to write effectively;
- * willingness to accept established essay scoring standards;
- * openness to the concepts and principles of holistic scoring; and,
- * a demonstrated ability to work well in group situations.

ESSAY READER CERTIFICATION:

To be certified as a GED essay reader, a qualified candidate:

- * attend a GEDTS-designed holistic scoring training session; and,
- * achieve an acceptable score on a set of reader certification papers provided by GEDTS.

Candidates who qualify for certification will be issued a GED Essay Reader Certificate by the state or province administrator. (Persons currently serving as GED teachers may not participate in the reading of GED examinee papers during an actual scoring session.)

CHIEF READER QUALIFICATIONS:

To apply for certification as a GED Chief Reader, a candidate must:

- * meet all essay reader qualifications;
- * have demonstrated leadership ability;
- * have strong communication skills; and,
- * have knowledge of holistic scoring procedures (participation in or leadership of scoring sessions preferred).

CHIEF READER CERTIFICATION:

To be certified as a GED Chief Reader, a candidate must be:

- * approved by the state or province administrator;
- * trained in holistic scoring procedures in accordance with GEDTS Chief Reader guidelines by attending a GEDTS Chief Reader training session;
- * willing to supervise GED holistic scoring sessions; and,
- * certified as a GED essay reader.

Candidates who qualify for certification will be issued a GED Chief Reader Certificate by the GED Testing Service.

GENERAL EDUCATIONAL TESTING SERVICE
WRITING SKILLS TEST, PART IIReader Objectivity Issues

Readers should be aware of and attempt to overcome the potential influence of personal bias on their ratings so that they can score each essay as objectively as possible. Bias can affect one's entire scoring pattern or may occur only in specific instances. An overall pattern such as the tendency to score consistently lower than other readers is generally easier to detect and correct than are idiosyncratic biases. A number of factors can bias or unknowingly influence a reader's perception of a paper in a positive or negative direction.

1. Some are characteristics of the paper itself such as:
 - a. handwriting;
 - b. lightness or darkness of writing;
 - c. neatness;
 - d. format (skipping lines, unusual margins, etc.);
 - e. presence or absence of a title; and
 - f. length of the response.
2. Other factors may be more difficult for a reader to identify since they are personal reactions. Perceptions of a paper's overall quality can be influenced by the readers:
 - a. expectations for writing performance that are developmentally or instructionally inappropriate for population being scored (e.g.,
 - b. personal reactions to the persona or tone of the composition (e.g., "What an egotist!"; "Sounds like a cute kid!"; "Another Erma Bombeck!"; etc.);
 - c. reactions to the content of the composition (e.g., "It doesn't say anything new."; "It's boring!"; "This paper is vulgar!"; "Not another paper about an elephant!"; etc.);
 - d. idiosyncratic preferences or prejudices in style or usage (e.g., the use of profanity; the use of "etc." or "that is");
 - e. deeply ingrained assessment standards from previous holistic scoring experience (e.g., this would be a 4 for ETS, our scoring guide is better, can I use it?);
 - f. assumptions about the writer or the testing conditions (e.g., "This examinee obviously has a lot of potential but ran out of time."; etc.). Score only what is on the paper.

GENERAL EDUCATIONAL DEVELOPMENT TESTING SERVICE
WRITING SKILLS TEST, PART II

Guidelines For Holistic Scoring

1. Read and Score. Read for an overall impression of the paper. Record your score immediately. Avoid the temptation to reread the paper to justify your score in terms of specific errors.
2. Red pens down. Holistic readers do not make marks on papers they score. There are two major reasons for this. First, there is not time to mark papers in this fashion. Each paper should be scored within a minute or two. Secondly, marking errors directs too much attention to one trait, while allowing the reader to ignore other positive features.
3. Do not change your score. Avoid second-guessing yourself. Make changes only if you made a recording error.
4. When in doubt, refer to the scoring guide criteria. Do not compare a paper to another paper or order papers in terms of quality. Instead, refer to the scale and your internalized concept of the score points.
5. Be aware that there can be a range of papers within a score point. Just as you have students making A's with 100 points and students making A's with 93 points, there are "high", "medium", and "low" papers within each score point.
6. Avoid the tendency to adjust the number of 6's, 5's, 3's, 2's, and 1's in a given set of papers. You may have an entire packet of 2's or an entire packet of 1's.
7. Be aware of and avoid potential biasing factors. (See Reader Objectivity Issues)
8. Score only what is on the paper. Avoid making assumptions about the writer's potential, why the writer omitted an idea, and so forth.
9. This is only a 45 minute test. Remember, examinees had only 45 minutes to read the prompt, to write a first draft, and to write the final draft. Do not compare the finished product with your writing assignments that students have at least a week to prepare.

BEST COPY AVAILABLE

Report on Essay Reading Site Certification/Monitoring

Site: EICD

READER REPORT

Total	Number of Essays	(%) Discepanant Scores		Holistic Score Points					
		w/GED	w/site	1	2	3	4	5	6
Site	120	6.7	15.0	7.1	32.5	28.3	19.2	8.8	4.2
GEDTS	120	0.0	0.0	3.3	29.2	45.8	15.0	5.0	1.7
Readers									
001	18	11.1	0.0	11.1	11.1	38.9	22.2	11.1	5.6
GEDTS	18	0.0	0.0	5.6	11.1	61.1	11.1	5.6	5.6
002	40	0.0	2.5	2.5	35.0	40.0	12.5	10.0	0.0
GEDTS	40	0.0	0.0	2.5	25.0	50.0	15.0	5.0	2.5
003	40	2.5	5.0	2.5	32.5	40.0	17.5	5.0	2.5
GEDTS	40	0.0	0.0	2.5	27.5	47.5	15.0	5.0	2.5
004	40	17.5	32.5	2.5	22.5	30.0	20.0	15.0	10.0
GEDTS	40	0.0	0.0	5.0	30.0	45.0	15.0	5.0	0.0
005	40	2.5	20.0	10.0	35.0	20.0	27.5	5.0	2.5
GEDTS	40	0.0	0.0	2.5	27.5	47.5	15.0	5.0	2.5
006	40	12.5	20.0	17.5	27.5	20.0	20.0	10.0	5.0
GEDTS	40	0.0	0.0	5.0	32.5	42.5	15.0	5.0	0.0
007	40	15.0	10.0	7.5	42.5	20.0	17.5	7.5	5.0
GEDTS	40	0.0	0.0	2.5	32.5	42.5	15.0	5.0	2.5

Evaluation Summary

(See attached interpretation of the site certification/monitoring results in relation to GEDTS standards.)

Report on Essay Reading Site Certification/Monitoring

Site: EICD

SCALE STABILITY (measured on 2 to 12 point scale)

Certification Criteria

Percent agreement with GEDTS..... 93.3%
 (percent of essays scored within 2 points of GEDTS score)

Percent of total essay scores

the same as GEDTS..... 26.7%
 within 1 point of GEDTS score..... 45.8%
 within 2 points of GEDTS score..... 20.8%

differing by more than 2 points from the GEDTS score. 6.7%

Correlation between GEDTS and Scoring Site..... .78

Reading Reliability	Site	GEDTS
Number of Essays Scored	120	120
Mean Total Score	6.02	5.88
Standard Deviation	2.20	1.93
Rater Agreement (1)		
Percent of essays with reader scores that are equal.....	45.0%	100%
differ by 1 point (contiguous).....	40.0%	
differ by 2 or more points (discrepant)	15.0%	
Correlation between raters	.63	1.0
Reading Reliability (2)	.88	

-
- (1) Rater agreement is based on the six-point holistic scale used by each reader.
- (2) Since the Essay Score is on the 2 to 12 point scale, correlations reported as "Correlations between Raters" have been stepped-up using the Spearman-Brown formula in order to represent the expected "Reading Reliability".

Appendix L

Essay Scores

Distributions of Resolved Scores(Rounded) - NOT Topic G or I

Kruskal-Wallis 1-way ANOVA

RESOLVD Resolved Scores (Rounded)
 NTOPIC Essay Topics of Interest

Mean Rank	Cases			
1074.83	620	NTOPIC =	2	Essay Topic B.
1067.83	611	NTOPIC =	3	Essay Topic C.
1107.80	315	NTOPIC =	6	Essay Topic F.
1130.68	309	NTOPIC =	10	Essay Topic J.
1023.49	301	NTOPIC =	14	Essay Topic N.

	2156	Total		

CASES	Chi-Square	Significance	Corrected for Ties Chi-Square	Significance
2156	5.4204	.2468	5.5047	.2393

Essay Scores and Multiple-Choice Standard Scores

(SCORES ARE BY TOPIC OF INTEREST)

VARIABLE	N	MEAN	STD DEV	SUM	MINIMUM	MAXIMUM
UNRES_B	264	6.22727273	2.35157389	1644.00000000	2.00000000	12.00000000
UNRES_C	289	5.90657439	2.30072264	1707.00000000	2.00000000	12.00000000
UNRES_F	124	6.12096774	2.31500142	759.00000000	2.00000000	11.00000000
UNRES_G	117	5.72649573	2.34750025	670.00000000	2.00000000	12.00000000
UNRES_I	130	6.41538462	2.57833735	834.00000000	2.00000000	12.00000000
UNRES_J	118	6.02542373	2.31480506	711.00000000	2.00000000	12.00000000
UNRES_N	136	5.91911765	2.35090258	805.00000000	2.00000000	12.00000000
STDSC_B	267	50.76779026	10.89933484	13555.00000000	20.00000000	80.00000000
STDSC_C	299	50.76588629	10.49282466	15179.00000000	21.00000000	80.00000000
STDSC_F	125	51.14400000	9.65760940	6393.00000000	20.00000000	80.00000000
STDSC_G	128	48.96093750	11.91267594	6267.00000000	20.00000000	77.00000000
STDSC_I	130	51.00000000	10.64144304	6630.00000000	20.00000000	80.00000000
STDSC_J	120	50.22500000	10.14471962	6027.00000000	20.00000000	73.00000000
STDSC_N	137	50.32116788	10.60031514	6894.00000000	20.00000000	80.00000000

PEARSON CORRELATION COEFFICIENTS / PROB / (N) UNDER H0:RHO=0 / NUMBER OF OBSERVATIONS

UNRES_B UNRES_C UNRES_F UNRES_G UNRES_I UNRES_J UNRES_N

STDSC_B	0.60026 0.0001 264	0	0	0	0	0
STDSC_C	0.58577 0.0001 289	0	0	0	0	0
STDSC_F	0.60049 0.0001 124	0	0	0	0	0
STDSC_G	0.48383 0.0001 117	0	0	0	0	0
STDSC_I	0.56337 0.0001 130	0	0	0	0	0
STDSC_J	0.56621 0.0001 118	0	0	0	0	0
STDSC_N	0.58967 0.0001 136	0	0	0	0	0

Multiple-Choice Standard Scores for persons taking a given topic

$$H_0: r_1 = r_2 = r_3 \dots r_n$$

$$\chi^2 = 2.59$$

$$df = 6$$

$$p = .85$$

H_0 is not rejected

Appendix L

DESCRIPTIVE DATA FOR STANDARD SCORES
ESSAYS BY TOPICS OF INTEREST
SIGNIFICANCE TEST AGAINST M.C. STD SCORES
ANALYSIS FOR VARIABLE STDScore CLASSIFIED BY VARIABLE MTopic

LEVEL	N	AVERAGE SCORES WERE USED FOR TIES		STD DEV UNDER MO	MEAN SCORE
		SUM OF SCORES	EXPECTED UNDER MO		
ESSAY TOPIC B.	267	144278.50	144046.50	4410.43	540.37
ESSAY TOPIC C.	299	160794.50	161310.50	4574.24	537.77
ESSAY TOPIC F.	125	70162.00	67437.50	3271.27	561.30
ESSAY TOPIC I.	130	71537.00	70135.00	3327.29	550.28
ESSAY TOPIC J.	120	62655.50	64740.00	3213.57	522.13
ESSAY TOPIC N.	137	72153.50	73911.50	3403.06	526.67

KRUSKAL-WALLIS TEST (CHI-SQUARE APPROXIMATION)
CHISQ= 1.39 DF= 5 PROB > CHISQ=0.9257
ESSAYS BY TOPICS OF INTEREST

Test of Multiple-Choice standard scores for groups taking various essay topics

DESCRIPTIVE DATA FOR STANDARD SCORES

VARIABLE	N	MEAN	STANDARD DEVIATION	SKEWNESS	KURTOSIS	MINIMUM VALUE	MAXIMUM VALUE	VARIANCE
----- TOPIC-B -----								
STDScore	267	50.76779026	10.89933484	0.05794108	-0.00288021	20.00000000	80.00000000	118.79549999
----- TOPIC-C -----								
STDScore	299	50.76588629	10.49282466	0.20173361	-0.21438413	21.00000000	80.00000000	110.09936926
----- TOPIC-F -----								
STDScore	125	51.44000000	9.65760940	-0.26344643	0.94250263	20.00000000	80.00000000	93.26941935
----- TOPIC-I -----								
STDScore	130	51.00000000	10.64144304	-0.03680942	0.47767348	20.00000000	80.00000000	113.24031008
----- TOPIC-J -----								
STDScore	120	50.22500000	10.14471962	0.10831512	0.14796662	20.00000000	73.00000000	102.91533613
----- TOPIC-N -----								
STDScore	137	50.32116788	10.60031514	0.15033303	0.85925780	20.00000000	80.00000000	112.36668098

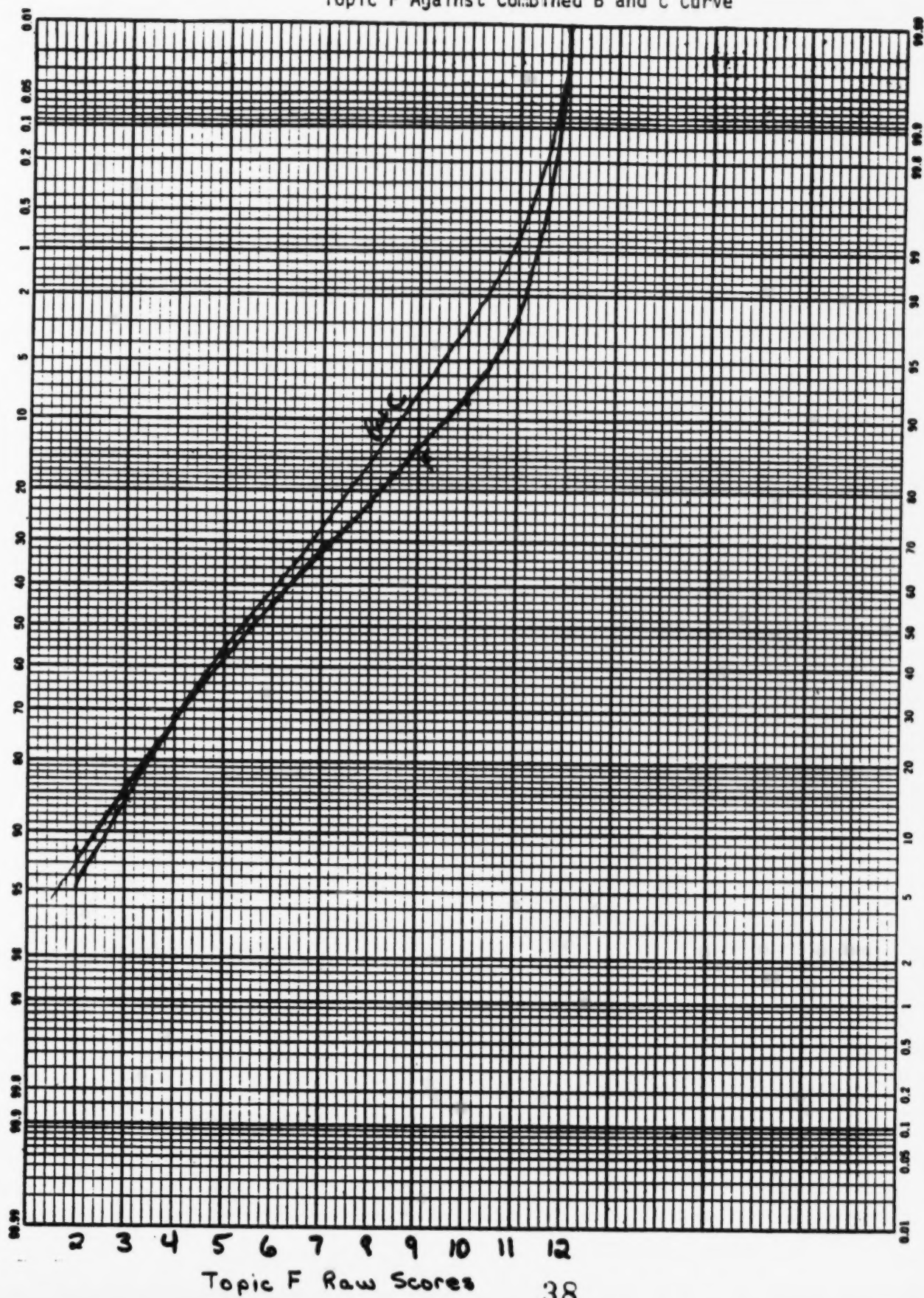
Appendix M

Essays (Topic F) - Resolved Rounded Scores

RESOLVD Resolved Score

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	2.00	26	8.2	8.3	8.3
	3.00	26	8.2	8.3	16.5
	4.00	38	11.9	12.1	28.6
	5.00	47	14.7	14.9	43.5
	6.00	38	11.9	12.1	55.6
	7.00	41	12.9	13.0	68.6
	8.00	28	8.8	8.9	77.5
	9.00	30	9.4	9.5	87.0
	10.00	15	4.7	4.8	91.7
	11.00	17	5.3	5.4	97.1
	12.00	9	2.8	2.9	100.0
	.	4	1.3	MISSING	
	TOTAL	319	100.0	100.0	
Valid Cases	315	Missing Cases	4		

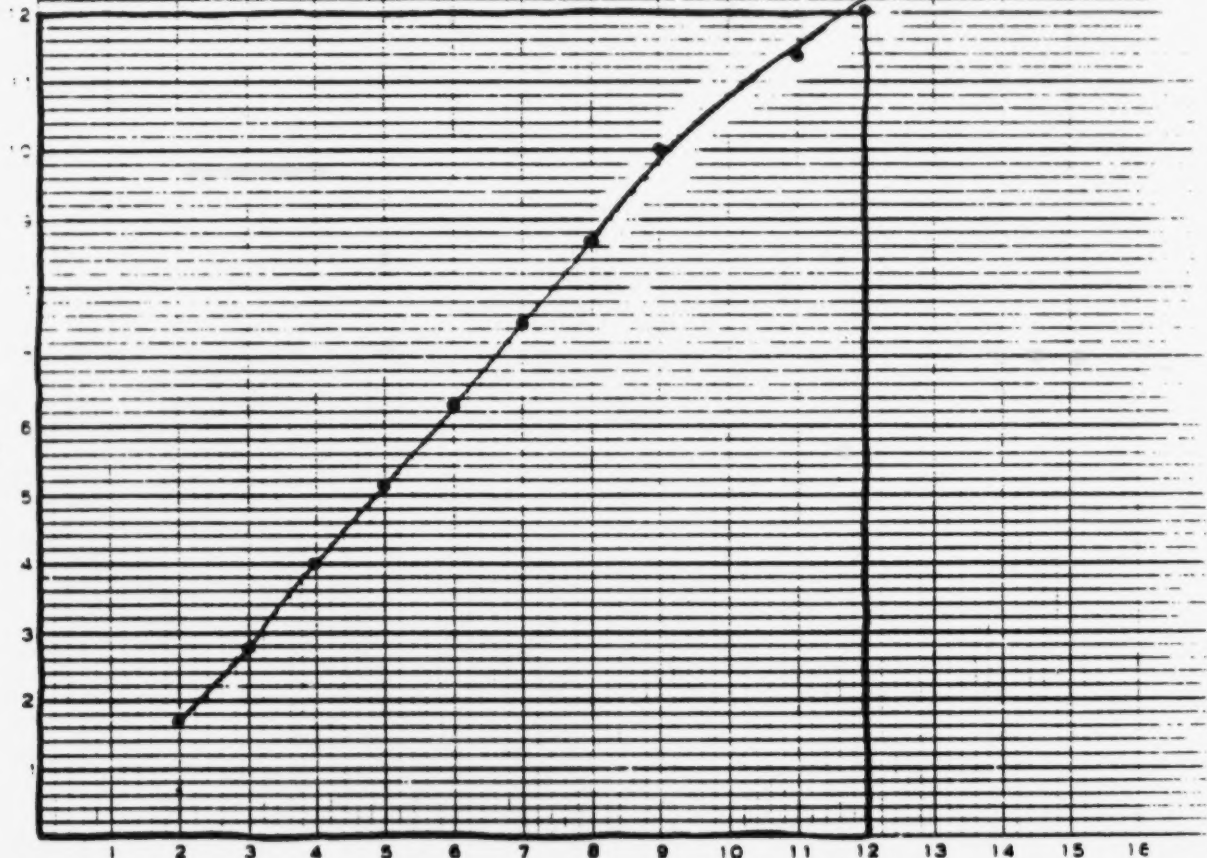
Topic F Against Combined B and C Curve



Topic F against B + C

Raw score	B+C Standard score	F Standard score
2	32-31	33
3	36	37
4	41	41
5	46	45
6	50	49
7	55	53
8	59-60	57
9	64-65	61
10	69	65
11	73-74	70
12	78-79	76

Topic
F
Score
Freq.



Topics B + C Score Frequencies

Weighted Composite Standard Score Formulas

In each of the following formulas, CS equals the weighted composite score, X equals the multiple choice standard score determined by the usual method (i.e., given the form of the multiple choice test, convert the raw score to a standard score on the 20 to 80 scale); Y equals the essay "raw score" (the score resulting from adding the two independent 1 to 6 reader scores resulting in the 2 to 12 scale).

Before the formulas below are applied, note that any two reader scores differing by more than one point for a given essay should be "resolved" by a third reading. The three reader scores for the essay must then be averaged, multiplied by 2, and rounded (when necessary) to the nearest whole number. (The rounding rule assumed in all computations is: .5 and above goes up; less than .5 goes down.)

The following two conditions must be applied once the composite score is calculated and rounded to the nearest whole number.

If CS < 20 then CS = 20
If CS > 80 then CS = 80

Formulas:

If topic equals B weighted composite score formula is:

$$CS = 0.7114 (X) + 0.4002 [4.4868 (Y) + 23.4592] - 5.6563$$

If topic equals C weighted composite score formula is:

$$CS = 0.7114 (X) + 0.4002 [4.6447 (Y) + 22.7628] - 5.6563$$

If topic equals F weighted composite score formula is:

$$CS = 0.7114 (X) + 0.4002 [3.9156 (Y) + 26.1840] - 5.6563$$

If topic equals J weighted composite score formula is:

$$CS = 0.7114 (X) + 0.4002 [4.4980 (Y) + 22.5327] - 5.6563$$

If topic equals N weighted composite score formula is:

$$CS = 0.7114 (X) + 0.4002 [4.2180 (Y) + 25.7816] - 5.6563$$
